

"Express Mail" Mailing Label No. EL 563 649 793 US

Atty Docket No. ROCH-002
Client No. R0066A

Date of Deposit: May 9, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.

Beth Berg

Typed or Printed Name of Person Mailing Paper or Fee



Signature of Person Mailing Paper or Fee

PATENT APPLICATION

FOR

HUMAN AGGRECANASE AND NUCLEIC ACID COMPOSITIONS ENCODING THE
SAME

Bret Field
Registration No. 37,620
BOZICEVIC, FIELD & FRANCIS LLP
200 Middlefield Road, Suite 200
Menlo Park, CA 94025

F:\DOCUMENT\ROCH (roche)\002\Application Cover Sheet.wpd

5

10

HUMAN AGGRECANASE AND NUCLEIC ACID COMPOSITIONS ENCODING THE SAME

CROSS REFERENCE TO RELATED APPLICATIONS

Pursuant to 35 U.S.C. § 119 (e), this application claims priority to the filing date of the
15 United States Provisional Patent Application Serial No. 60/133,343 filed May 10, 1999, the
disclosure of which is herein incorporated by reference.

INTRODUCTION

Field of the Invention

20 The field of the invention is proteases, particularly proteases that cleave aggrecan.

Background of the Invention

Cartilage matrix structure as dry weight of the tissue is made up of 70% collagen and 20-
30% proteoglycans. The proteoglycan component confers mechanical flexibility to load bearing
25 tissues and imparts viscoelastic properties to cartilage. Its loss leads to rapid structural damage as is
seen most frequently in arthritic joint diseases and joint injury.

Aggrecan is a major cartilage proteoglycan. Aggrecan is a large protein of 210 kDa and has
three globular domains: G1, G2, and G3. The G1 and G2 domains of the protein are closer to the
amino terminus of the protein and their intervening interglobular domain has sites that are
30 proteolytically sensitive. The region between G2 and G3 is heavily glycosylated and connected to
oligosaccharides and glycosaminoglycans (GAGs) to form the mature proteoglycan. In arthritic

cartilage, core protein fragments of 55 kDa are observed and believed to be the result of cleavage of the core protein in the G1 and G2 interglobular domain between asparagine 341 and phenylalanine 342. This cleavage can be made by many matrix metalloproteinases e.g. MMP-1, -2, -3, -7, -8, -9, and -13. In addition, 60 kDa aggrecan fragments with a -COOH terminus of glutamic acid are also
5 identified and are indicative of a cleavage site between glutamic acid 373 and alanine 374. Matrix metalloproteinase are unable to cleave at this site. The unique endopeptidase activity responsible for this cleavage has been termed "aggrecanase."

The G1 domain of the core protein forms a stable ternary complex by binding to hyaluronic acid and link proteins in the matrix. Any enzymatic cleavage in this region destabilizes the cartilage
10 matrix structure, leads to the loss of the major proteoglycan aggrecan and exposes type II collagen to collagenases, causing cartilage loss and the consequent development of joint disease. Since a variety of anti-arthritic drugs do not target aggrecanase and are incapable of blocking cleavage of aggrecan, the aggrecanase site plays a key role in the proteolytic degradation of aggrecan.

As such, aggrecanase is considered to be an important drug target for arthritis. Aggrecan
15 fragments released into the synovial fluid are the primary detectable events in the development of rheumatoid- and osteo- arthritis. Search for this protease has been intense. Despite these intense discovery efforts, identification of human aggrecanase has remained elusive.

As such, there is much interest in the identification of human aggrecanase, as well as the gene encoding this activity.

20 Relevant Literature

U.S. Patents of interest include: 5,872,209 and 5,427,954. PCT publications of interest include: WO 99/09000; WO 98/55643; WO 98/51665; and WO 97/18207.

Other references of interest include: Vankemmelbeke et al., "Coincubation of bovine synovial or capsular tissue with cartilage generates a soluble 'Aggrecanase' activity," Biochem
25 Biophys Res Commun (1999 Feb 24) 255(3):686-91; Arner et al., "Generation and Characterization of Aggrecanase. A soluble, cartilage-derived aggrecan-degrading activity," J Biol Chem (1999 Mar 5) 274(10):6594-6601; Billington et al., "An aggrecan-degrading activity associated with chondrocyte membranes," Biochem J (1998 Nov 15) 336 (Pt 1):207-12; Hughes et al., "Differential expression of aggrecanase and matrix metalloproteinase activity in chondrocytes
30 isolated from bovine and porcine articular cartilage," J Biol Chem (1998 Nov 13) 273(46):30576-82; Sandy et al., "Chondrocyte-mediated catabolism of aggrecan: aggrecanase-dependent cleavage

induced by interleukin-1 or retinoic acid can be inhibited by glucosamine," Biochem J (1998 Oct 1) 335 (Pt 1):59-66; Arner et al., "Cytokine-induced cartilage proteoglycan degradation is mediated by aggrecanase," Osteoarthritis Cartilage (1998 May) 6(3):214-28; Ilic et al., "Characterization of aggrecan retained and lost from the extracellular matrix of articular cartilage. Involvement of carboxyl-terminal processing in the catabolism of aggrecan," J Biol Chem (1998 Jul 10) 273(28):17451-8; and Buttner et al., "Membrane type 1 matrix metalloproteinase (MT1-MMP) cleaves the recombinant aggrecan substrate rAgg1 mut at the 'aggrecanase' and the MMP sites. Characterization of MT1-MMP catabolic activities on the interglobular domain of aggrecan," Biochem J (1998 Jul 1) 333 (Pt 1):159-65.

10

SUMMARY OF THE INVENTION

Human aggrecanase and polypeptides related thereto, as well as nucleic acid compositions encoding the same, are provided. The subject polypeptide and nucleic acid compositions find use in a variety of applications, including research, diagnostic, and therapeutic agent screening applications, as well as in treatment therapies. Also provided are methods of treating disease conditions associated with aggrecanase activity, e.g. conditions characterized by the presence of aggrecan cleavage products, such as rheumatoid- and osteo-arthritis.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 provides the nucleic acid and amino acid sequence of human aggrecanase.

DETAILED DESCRIPTION OF THE INVENTION

A novel human enzyme having aggrecan cleavage activity (i.e. human aggrecanase or ADAMTS-1) and polypeptides related thereto, as well as nucleic acid compositions encoding the same, are provided. The subject polypeptide and/or nucleic acid compositions find use in a variety of different applications, including research, diagnostic, and therapeutic agent screening/discovery/preparation applications. Also provided are methods of treating disease conditions associated with aggrecanase function, e.g. diseases characterized by the presence of aggrecan cleavage products, such as rheumatoid- and osteo-arthritis.

Before the subject invention is further described, it is to be understood that the invention is not limited to the particular embodiments of the invention described below, as variations of the particular embodiments may be made and still fall within the scope of the appended claims. It is also to be understood that the terminology employed is for the purpose of describing particular
5 embodiments, and is not intended to be limiting. Instead, the scope of the present invention will be established by the appended claims.

In this specification and the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Unless defined otherwise, all
10 technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs.

POLYPEPTIDE COMPOSITIONS

15 A novel human enzyme having aggrecanase activity (i.e. human aggrecanase or ADAMTS-1), as well as polypeptide compositions related thereto, are provided. The term polypeptide composition as used herein refers to both the full length human protein, as well as portions or fragments thereof. Also included in this term are variations of the naturally occurring human
20 protein, where such variations are homologous or substantially similar to the naturally occurring protein, as described in greater detail below. In the following description of the subject invention, the term "aggrecanase" is used to refer to the wild type human aggrecanase molecule of the subject invention.

The aggrecanase protein of the subject invention is an enzyme, particularly a proteinase and
25 more particularly a metalloproteinase. The subject human aggrecanase is characterized by having aggrecanase activity. As such, the subject aggrecanase is capable of cleaving aggrecan in an interglobular domain, particularly between the G1 and G2 domains, and more particularly at the Glu³⁷³-Ala³⁷⁴ bond of human aggrecan, to produce a cleavage product having an N-terminal sequence of ARGSVIL.

30 Human aggrecanase has an amino acid sequence as shown in Fig. 1 and identified as SEQ ID NO:02. Human aggrecanase has a molecular weight based on its amino acid sequence of about

100 kDa. The true molecular weight of human aggrecanase may vary due to glycosylation and/or other postranslational modifications. As such, the actual molecular weight of human aggrecanase is likely to be in the range from about 80 to 100 kDa, usually from about 85 to 95 kDa.

Aggrecanase homologs or proteins (or fragments thereof) that vary in sequence from the wild type sequence of the subject invention are also provided. By homolog is meant a protein having at least about 35 %, usually at least about 40% and more usually at least about 60 % amino acid sequence identity to the human aggrecanase protein of the subject invention, as determined using MegAlign, DNASTar (1998) clustal algorithm as described in D. G. Higgins and P.M. Sharp, "Fast and Sensitive multiple Sequence Alignments on a Microcomputer," (1989) CABIOS, 5: 151-153. (Parameters used are ktuple 1, gpa penalty 3, window, 5 and diagonals saved 5).

Also provided are aggrecanase proteins that are substantially identical to the hu aggrecanase protein, where by substantially identical is meant that the protein has an amino acid sequence identity to the sequence of aggrecanase of at least about 60%, usually at least about 65% and more usually at least about 70 %.

The proteins of the subject invention are present in a non-naturally occurring environment, *e.g.* are separated from their naturally occurring environment. In certain embodiments, the subject proteins are present in a composition that is enriched for the subject protein as compared to its naturally occurring environment. For example, purified aggrecanase is provided, where by purified is meant that the aggrecanase enzyme is present in a composition that is substantially free of non-aggrecanase proteins, where by substantially free is meant that less than 90 %, usually less than 60 % and more usually less than 50 % of the composition is made up of non-aggrecanase proteins. The proteins of the subject invention may also be present as an isolate, by which is meant that the protein is substantially free of other proteins and other naturally occurring biologic molecules, such as oligosaccharides, polynucleotides and fragments thereof, and the like, where the term "substantially free" in this instance means that less than 70 %, usually less than 60% and more usually less than 50 % of the composition containing the isolated protein is some other naturally occurring biological molecule. In certain embodiments, the proteins are present in substantially pure form, where by "substantially pure form" is meant at least 95%, usually at least 97% and more usually at least 99% pure.

In addition to the naturally occurring proteins, polypeptides which vary from the naturally occurring proteins are also provided, *e.g.* aggrecanase polypeptides. By aggrecanase polypeptide is

meant an amino acid sequence encoded by an open reading frame (ORF) of the gene encoding aggrecanase, described in greater detail below, including the full length aggrecanase protein and fragments thereof, particularly biologically active fragments and/or fragments corresponding to functional domains, e.g. transmembrane domain, and the like; and including fusions of the subject polypeptides to other proteins or parts thereof. Fragments of interest will typically be at least about 10 aa in length, usually at least about 50 aa in length, and may be as long as 300 aa in length or longer, but will usually not exceed about 1000 aa in length, where the fragment will have a stretch of amino acids that is identical to the subject protein of at least about 10 aa, and usually at least about 15 aa, and in many embodiments at least about 50 aa in length.

The subject proteins and polypeptides may be obtained from naturally occurring sources or synthetically produced. For example, human aggrecanase may be derived from biological sources which express aggrecanase, such as synoviocytes, chondrocytes, cartilage and the like. The subject proteins may also be derived from synthetic means, e.g. by expressing a recombinant gene encoding protein of interest in a suitable host, as described in greater detail below. Any convenient protein purification procedures may be employed, where suitable protein purification methodologies are described in Guide to Protein Purification, (Deutscher ed.) (Academic Press, 1990). For example, a lysate may be prepared from the original source, e.g. chondrocytes or the expression host, and purified using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, and the like.

NUCLEIC ACID COMPOSITIONS

Also provided are nucleic acid compositions encoding aggrecanase proteins or fragments thereof, as well as the aggrecanase homologues of the present invention. By aggrecanase nucleic acid composition is meant a composition comprising a sequence of DNA having an open reading frame that encodes aggrecanase, i.e. an *aggrecanase* gene, and is capable, under appropriate conditions, of being expressed as aggrecanase. Also encompassed in this term are nucleic acids that are homologous or substantially similar or identical to the nucleic acids encoding aggrecanase proteins. Thus, the subject invention provides genes encoding the human aggrecanase of the subject invention and homologs thereof. The human aggrecanase gene has the nucleic acid sequence shown in Fig. 1 and identified as SEQ ID NO:01, *infra*.

The source of homologous genes may be any species, e.g., primate species, particularly human; rodents, such as rats and mice, canines, felines, bovines, ovines, equines, yeast, nematodes, etc. Between mammalian species, e.g., human and mouse, homologs have substantial sequence similarity, e.g. at least 75% sequence identity, usually at least 90%, more usually at least 95% between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, etc. A reference sequence will usually be at least about 18 nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as BLAST, described in Altschul *et al.* (1990), *J. Mol. Biol.* 215:403-10 (using default settings, i.e. parameters $w=4$ and $T=17$). The sequences provided herein are essential for recognizing aggrecanase-related and homologous proteins, and the nucleic acids encoding the same, in database searches.

Nucleic acids encoding the aggrecanase protein and aggrecanase polypeptides of the subject invention may be cDNA or genomic DNA or a fragment thereof. The term “aggrecanase gene” shall be intended to mean the open reading frame encoding specific aggrecanase proteins and polypeptides, and aggrecanase introns, as well as adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 20 kb beyond the coding region, but possibly further in either direction. The gene may be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome.

The term “cDNA” as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 5' and 3' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding an aggrecanase protein.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It may further include 5' and 3' un-translated regions found in the mature mRNA. It may further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' or 3' end of the transcribed region. The genomic DNA may be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal

sequence. The genomic DNA flanking the coding region, either 3' or 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue and stage specific expression.

5 The nucleic acid compositions of the subject invention may encode all or a part of the subject aggrecanase protein. Double or single stranded fragments may be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* For the most part, DNA fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and may be at least about 50 nt.

10 The *aggrecanase* genes are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the DNA will be obtained substantially free of other nucleic acid sequences that do not include an *aggrecanase* sequence or fragment thereof, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *i.e.* flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

15

PREPARATION OF AGGREGANASE POLYPEPTIDES

In addition to the plurality of uses described in greater detail in following sections, the subject nucleic acid compositions find use in the preparation of all or a portion of the aggrecanase polypeptides, as described above. The provided polynucleotide (*e.g.*, a polynucleotide having a sequence of SEQ ID NO:01, the corresponding cDNA, or the full-length gene is used to express a partial or complete gene product. Constructs of polynucleotides having a sequences of "SEQ ID NO:01 can be generated synthetically. Alternatively, single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides is described by, *e.g.*, Stemmer *et al.*, 20 *Gene (Amsterdam)* (1995) 164(1):49-53. In this method, assembly PCR (the synthesis of long DNA sequences from large numbers of oligodeoxyribonucleotides (oligos)) is described. The method is derived from DNA shuffling (Stemmer, *Nature* (1994) 370:389-391), and does not rely on DNA ligase, but instead relies on DNA polymerase to build increasingly longer DNA fragments during the assembly process. Appropriate polynucleotide constructs are purified using standard 25 recombinant DNA techniques as described in, for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY, and under 30

current regulations described in United States Dept. of HHS, National Institute of Health (NIH) Guidelines for Recombinant DNA Research.

Polynucleotide molecules comprising a polynucleotide sequence provided herein are propagated by placing the molecule in a vector. Viral and non-viral vectors are used, including plasmids. The choice of plasmid will depend on the type of cell in which propagation is desired and the purpose of propagation. Certain vectors are useful for amplifying and making large amounts of the desired DNA sequence. Other vectors are suitable for expression in cells in culture. Still other vectors are suitable for transfer and expression in cells in a whole animal or person. The choice of appropriate vector is well within the skill of the art. Many such vectors are available commercially. The partial or full-length polynucleotide is inserted into a vector typically by means of DNA ligase attachment to a cleaved restriction enzyme site in the vector. Alternatively, the desired nucleotide sequence can be inserted by homologous recombination in vivo. Typically this is accomplished by attaching regions of homology to the vector on the flanks of the desired nucleotide sequence. Regions of homology are added by ligation of oligonucleotides, or by polymerase chain reaction using primers comprising both the region of homology and a portion of the desired nucleotide sequence, for example.

For expression, an expression cassette or system may be employed. The gene product encoded by a polynucleotide of the invention is expressed in any convenient expression system, including, for example, bacterial, yeast, insect, amphibian and mammalian systems. Suitable vectors and host cells are described in U.S. Patent No. 5,654,173. In the expression vector, an aggrecanase encoding polynucleotide, e.g. as set forth in SEQ ID NO:01, is linked to a regulatory sequence as appropriate to obtain the desired expression properties. These can include promoters (attached either at the 5' end of the sense strand or at the 3' end of the antisense strand), enhancers, terminators, operators, repressors, and inducers. The promoters can be regulated or constitutive. In some situations it may be desirable to use conditionally active promoters, such as tissue-specific or developmental stage-specific promoters. These are linked to the desired nucleotide sequence using the techniques described above for linkage to vectors. Any techniques known in the art can be used. In other words, the expression vector will provide a transcriptional and translational initiation region, which may be inducible or constitutive, where the coding region is operably linked under the transcriptional control of the transcriptional initiation region, and a transcriptional and translational

termination region. These control regions may be native to the *aggrecanase* gene, or may be derived from exogenous sources.

Expression vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences encoding heterologous proteins. A selectable marker operative in the expression host may be present. Expression vectors may be used for the production of fusion proteins, where the exogenous fusion peptide provides additional functionality, i.e. increased protein synthesis, stability, reactivity with defined antisera, an enzyme marker, e.g. β -galactosidase, etc.

Expression cassettes may be prepared comprising a transcription initiation region, the gene or fragment thereof, and a transcriptional termination region. Of particular interest is the use of sequences that allow for the expression of functional epitopes or domains, usually at least about 8 amino acids in length, more usually at least about 15 amino acids in length, to about 25 amino acids, and up to the complete open reading frame of the gene. After introduction of the DNA, the cells containing the construct may be selected by means of a selectable marker, the cells expanded and then used for expression.

Aggrecanase proteins and polypeptides may be expressed in prokaryotes or eukaryotes in accordance with conventional ways, depending upon the purpose for expression. For large scale production of the protein, a unicellular organism, such as *E. coli*, *B. subtilis*, *S. cerevisiae*, insect cells in combination with baculovirus vectors, or cells of a higher organism such as vertebrates, particularly mammals, e.g. COS 7 cells, HEK 293, CHO, *Xenopus* Oocytes, etc., may be used as the expression host cells. In some situations, it is desirable to express the *aggrecanase* gene in eukaryotic cells, where the aggrecanase protein will benefit from native folding and post-translational modifications. Small peptides can also be synthesized in the laboratory. Polypeptides that are subsets of the complete aggrecanase sequence may be used to identify and investigate parts of the protein important for function.

Specific expression systems of interest include bacterial, yeast, insect cell and mammalian cell derived expression systems. Representative systems from each of these categories is are provided below:

Bacteria. Expression systems in bacteria include those described in Chang *et al.*, *Nature* (1978) 275:615; Goeddel *et al.*, *Nature* (1979) 281:544; Goeddel *et al.*, *Nucleic Acids Res.* (1980)

8:4057; EP 0 036,776; U.S. Patent No. 4,551,433; DeBoer *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1983) 80:21-25; and Siebenlist *et al.*, *Cell* (1980) 20:269.

Yeast. Expression systems in yeast include those described in Hinnen *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1978) 75:1929; Ito *et al.*, *J. Bacteriol.* (1983) 153:163; Kurtz *et al.*, *Mol. Cell. Biol.* (1986) 6:142; Kunze *et al.*, *J. Basic Microbiol.* (1985) 25:141; Gleeson *et al.*, *J. Gen. Microbiol.* (1986) 132:3459; Roggenkamp *et al.*, *Mol. Gen. Genet.* (1986) 202:302; Das *et al.*, *J. Bacteriol.* (1984) 158:1165; De Louvencourt *et al.*, *J. Bacteriol.* (1983) 154:737; Van den Berg *et al.*, *Bio/Technology* (1990) 8:135; Kunze *et al.*, *J. Basic Microbiol.* (1985) 25:141; Cregg *et al.*, *Mol. Cell. Biol.* (1985) 5:3376; U.S. Patent Nos. 4,837,148 and 4,929,555; Beach and Nurse, *Nature* (1981) 300:706; Davidow *et al.*, *Curr. Genet.* (1985) 10:380; Gaillardin *et al.*, *Curr. Genet.* (1985) 10:49; Ballance *et al.*, *Biochem. Biophys. Res. Commun.* (1983) 112:284-289; Tilburn *et al.*, *Gene* (1983) 26:205-221; Yelton *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1984) 81:1470-1474; Kelly and Hynes, *EMBO J.* (1985) 4:475479; EP 0 244,234; and WO 91/00357.

Insect Cells. Expression of heterologous genes in insects is accomplished as described in U.S. Patent No. 4,745,051; Friesen *et al.*, "The Regulation of Baculovirus Gene Expression", in: *The Molecular Biology Of Baculoviruses* (1986) (W. Doerfler, ed.); EP 0 127,839; EP 0 155,476; and Vlak *et al.*, *J. Gen. Virol.* (1988) 69:765-776; Miller *et al.*, *Ann. Rev. Microbiol.* (1988) 42:177; Carbonell *et al.*, *Gene* (1988) 73:409; Maeda *et al.*, *Nature* (1985) 315:592-594; Lebacqz-Verheyden *et al.*, *Mol. Cell. Biol.* (1988) 8:3129; Smith *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1985) 82:8844; Miyajima *et al.*, *Gene* (1987) 58:273; and Martin *et al.*, *DNA* (1988) 7:99. Numerous baculoviral strains and variants and corresponding permissive insect host cells from hosts are described in Luckow *et al.*, *Bio/Technology* (1988) 6:47-55, Miller *et al.*, *Generic Engineering* (1986) 8:277-279, and Maeda *et al.*, *Nature* (1985) 315:592-594.

Mammalian Cells. Mammalian expression is accomplished as described in Dijkema *et al.*, *EMBO J.* (1985) 4:761, Gorman *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1982) 79:6777, Boshart *et al.*, *Cell* (1985) 41:521 and U.S. Patent No. 4,399,216. Other features of mammalian expression are facilitated as described in Ham and Wallace, *Meth. Enz.* (1979) 58:44, Barnes and Sato, *Anal. Biochem.* (1980) 102:255, U.S. Patent Nos. 4,767,704, 4,657,866, 4,927,762, 4,560,655, WO 90/103430, WO 87/00195, and U.S. RE 30,985.

When any of the above host cells, or other appropriate host cells or organisms, are used to replicate and/or express the polynucleotides or nucleic acids of the invention, the resulting

replicated nucleic acid, RNA, expressed protein or polypeptide, is within the scope of the invention as a product of the host cell or organism. The product is recovered by any appropriate means known in the art.

Once the gene corresponding to a selected polynucleotide is identified, its expression can be regulated in the cell to which the gene is native. For example, an endogenous gene of a cell can be regulated by an exogenous regulatory sequence inserted into the genome of the cell at location sufficient to at least enhance expression of the gene in the cell. The regulatory sequence may be designed to integrate into the genome via homologous recombination, as disclosed in U.S. Patent Nos. 5,641,670 and 5,733,761, the disclosures of which are herein incorporated by reference, or may be designed to integrate into the genome via non-homologous recombination, as described in WO 99/15650, the disclosure of which is herein incorporated by reference. As such, also encompassed in the subject invention is the production of the subject aggrecanase proteins without manipulation of the encoding nucleic acid itself, but instead through integration of a regulatory sequence into the genome of cell that already includes a gene encoding the desired protein, as described in the above incorporated patent documents.

USES OF THE SUBJECT AGGREGANASE POLYPEPTIDE AND NUCLEIC ACID COMPOSITIONS

The subject polypeptide and nucleic acid compositions find use in a variety of different applications, including research, diagnostic, and therapeutic agent screening/discovery/ preparation applications, as well as in therapeutic compositions and methods employing the same.

RESEARCH APPLICATIONS

The subject nucleic acid compositions find use in a variety of research applications. Research applications of interest include: the identification of aggrecanase homologs; as a source of novel promoter elements; the identification of aggrecanase expression regulatory factors; as probes and primers in hybridization applications, e.g. PCR; the identification of expression patterns in biological specimens; the preparation of cell or animal models for aggrecanase function; the preparation of *in vitro* models for aggrecanase function; etc.

Homologs of the *aggrecanase* gene are identified by any of a number of methods. A fragment of the provided cDNA may be used as a hybridization probe against a cDNA library from

the target organism of interest, where low stringency conditions are used. The probe may be a large fragment, or one or more short degenerate primers. Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 6×SSC (0.9 M sodium chloride/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1×SSC (0.15 M sodium chloride/0.015 M sodium citrate). Sequence identity may be determined by hybridization under stringent conditions, for example, at 50°C or higher and 0.1×SSC (15 mM sodium chloride/0.15 mM sodium citrate). Nucleic acids having a region of substantial identity to the provided *aggrecanase* sequences, e.g. allelic variants, genetically altered versions of the gene, etc., bind to the provided *aggrecanase* sequences under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes.

The sequence of the 5' flanking region may be utilized for promoter elements, including enhancer binding sites, that provide for developmental regulation in tissues where *aggrecanase* is expressed. The tissue specific expression is useful for determining the pattern of expression, and for providing promoters that mimic the native pattern of expression. Naturally occurring polymorphisms in the promoter region are useful for determining natural variations in expression, particularly those that may be associated with disease.

Alternatively, mutations may be introduced into the promoter region to determine the effect of altering expression in experimentally defined systems. Methods for the identification of specific DNA motifs involved in the binding of transcriptional factors are known in the art, e.g. sequence similarity to known binding motifs, gel retardation studies, etc. For examples, see Blackwell *et al.* (1995), *Mol. Med.* 1:194-205; Mortlock *et al.* (1996), *Genome Res.* 6:327-33; and Joulin and Richard-Foy (1995), *Eur. J. Biochem.* 232:620-626.

The regulatory sequences may be used to identify *cis* acting sequences required for transcriptional or translational regulation of *aggrecanase* gene expression, especially in different tissues or stages of development, and to identify *cis* acting sequences and *trans*-acting factors that regulate or mediate *aggrecanase* gene expression. Such transcription or translational control regions may be operably linked to an *aggrecanase* gene in order to promote expression of wild type or altered *aggrecanase* or other proteins of interest in cultured cells, or in embryonic, fetal or adult tissues, and for gene therapy.

Small DNA fragments are useful as primers for PCR, hybridization screening probes, *etc.* Larger DNA fragments, *i.e.* greater than 100 nt are useful for production of the encoded polypeptide, as described in the previous section. For use in geometric amplification reactions, such as geometric PCR, a pair of primers will be used. The exact composition of the primer sequences is not critical to the invention, but for most applications the primers will hybridize to the subject sequence under stringent conditions, as known in the art. It is preferable to choose a pair of primers that will generate an amplification product of at least about 50 nt, preferably at least about 100 nt. Algorithms for the selection of primer sequences are generally known, and are available in commercial software packages. Amplification primers hybridize to complementary strands of DNA, and will prime towards each other.

The DNA may also be used to identify expression of the gene in a biological specimen. The manner in which one probes cells for the presence of particular nucleotide sequences, as genomic DNA or RNA, is well established in the literature. Briefly, DNA or mRNA is isolated from a cell sample. The mRNA may be amplified by RT-PCR, using reverse transcriptase to form a complementary DNA strand, followed by polymerase chain reaction amplification using primers specific for the subject DNA sequences. Alternatively, the mRNA sample is separated by gel electrophoresis, transferred to a suitable support, *e.g.* nitrocellulose, nylon, *etc.*, and then probed with a fragment of the subject DNA as a probe. Other techniques, such as oligonucleotide ligation assays, *in situ* hybridizations, and hybridization to DNA probes arrayed on a solid chip may also find use. Detection of mRNA hybridizing to the subject sequence is indicative of aggrecanase gene expression in the sample.

The sequence of an aggrecanase gene, including flanking promoter regions and coding regions, may be mutated in various ways known in the art to generate targeted changes in promoter strength, sequence of the encoded protein, *etc.* The DNA sequence or protein product of such a mutation will usually be substantially similar to the sequences provided herein, *i.e.* will differ by at least one nucleotide or amino acid, respectively, and may differ by at least two but not more than about ten nucleotides or amino acids. The sequence changes may be substitutions, insertions, deletions, or a combination thereof. Deletions may further include larger changes, such as deletions of a domain or exon. Other modifications of interest include epitope tagging, *e.g.* with the FLAG system, HA, *etc.* For studies of subcellular localization, fusion proteins with green fluorescent proteins (GFP) may be used.

Techniques for *in vitro* mutagenesis of cloned genes are known. Examples of protocols for site specific mutagenesis may be found in Gustin *et al.* (1993), *Biotechniques* 14:22; Barany (1985), *Gene* 37:111-23; Colicelli *et al.* (1985), *Mol. Gen. Genet.* 199:537-9; and Prentki *et al.* (1984), *Gene* 29:303-13. Methods for site specific mutagenesis can be found in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, CSH Press 1989, pp. 15.3-15.108; Weiner *et al.* (1993), *Gene* 126:35-41; Sayers *et al.* (1992), *Biotechniques* 13:592-6; Jones and Winistorfer (1992), *Biotechniques* 12:528-30; Barton *et al.* (1990), *Nucleic Acids Res* 18:7349-55; Marotti and Tomich (1989), *Gene Anal. Tech.* 6:67-70; and Zhu (1989), *Anal Biochem* 177:120-4. Such mutated genes may be used to study structure-function relationships of aggrecanase, or to alter properties of the protein that affect its function or regulation.

The subject nucleic acids can be used to generate transgenic, non-human animals or site specific gene modifications in cell lines. Transgenic animals may be made through homologous recombination, where the endogenous aggrecanase locus is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like.

The modified cells or animals are useful in the study of aggrecanase function and regulation. Of interest is the use of aggrecanase genes to construct transgenic animal models of aggrecanase related disease conditions, e.g. disease conditions associated with aggrecanase activity, such as arthritis. Thus, transgenic animal models of the subject invention include endogenous *aggrecanase* knockouts in which expression of endogenous aggrecanase is at least reduced if not eliminated, where such animals also typically express an aggrecanase peptide of the subject invention, e.g. the aggrecanase protein of the subject invention or a fragment thereof. Where a nucleic acid having a sequence found in the human aggrecanase gene is introduced, the introduced nucleic acid may be either a complete or partial sequence of the aggrecanase gene. A detectable marker, such as *lac Z* may be introduced into the aggrecanase locus, where upregulation of aggrecanase expression will result in an easily detected change in phenotype. One may also provide for expression of the aggrecanase gene or variants thereof in cells or tissues where it is not normally expressed, at levels not normally present in such cells or tissues.

DNA constructs for homologous recombination will comprise at least a portion of the aggrecanase gene of the subject invention, wherein the gene has the desired genetic modification(s), and includes regions of homology to the target locus. DNA constructs for random integration need

not include regions of homology to mediate recombination. Conveniently, markers for positive and negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art. For various techniques for transfecting mammalian cells, see Keown *et al.* (1990), *Meth. Enzymol.* **185**:527-537.

5 For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, *e.g.* mouse, rat, guinea pig, *etc.* Such cells are grown on an appropriate fibroblast-feeder layer or grown in the presence of leukemia inhibiting factor (LIF). When ES or embryonic cells have been transformed, they may be used to produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. 10 Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination or integration of the construct. Those colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old superovulated females. The ES cells are trypsinized, and the modified cells are injected into the 15 blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting offspring screened for the construct. By providing for a different phenotype of the blastocyst and the genetically modified cells, chimeric progeny can be readily detected.

The chimeric animals are screened for the presence of the modified gene and males and 20 females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogeneic or congenic grafts or transplants, or in *in vitro* culture. The transgenic animals may be any non-human mammal, such as laboratory animals, domestic animals, *etc.* The transgenic animals may be used in functional studies, drug screening, *etc.*, *e.g.* to determine the effect of a candidate drug on 25 aggrecanase activity.

DIAGNOSTIC APPLICATIONS

Also provided are methods of diagnosing disease states based on observed levels of 30 aggrecanase or the expression level of the aggrecanase gene in a biological sample of interest. Samples, as used herein, include biological fluids such as blood, cerebrospinal fluid, tears, saliva,

lymph, dialysis fluid, and the like; organ or tissue culture derived fluids; and fluids extracted from physiological tissues. Also included in the term are derivatives and fractions of such fluids. The cells may be dissociated, in the case of solid tissues, or tissue sections may be analyzed. Alternatively a lysate of the cells may be prepared.

5 A number of methods are available for determining the expression level of a gene or protein in a particular sample. Diagnosis may be performed by a number of methods to determine the absence or presence or altered amounts of normal or abnormal aggrecanase in a patient sample. For example, detection may utilize staining of cells or histological sections with labeled antibodies, performed in accordance with conventional methods. Cells are permeabilized to stain cytoplasmic
10 molecules. The antibodies of interest are added to the cell sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody may be labeled with radioisotopes, enzymes, fluorescers, chemilumescers, or other labels for direct detection. Alternatively, a second stage antibody or reagent is used to amplify the signal. Such reagents are well known in the art. For example, the primary antibody may be conjugated to biotin,
15 with horseradish peroxidase-conjugated avidin added as a second stage reagent. Alternatively, the secondary antibody conjugated to a fluorescent compound, *e.g.* fluorescein, rhodamine, Texas red, *etc.* Final detection uses a substrate that undergoes a color change in the presence of the peroxidase. The absence or presence of antibody binding may be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, *etc.*

20 Alternatively, one may focus on the expression of aggrecanase. Biochemical studies may be performed to determine whether a sequence polymorphism in an aggrecanase coding region or control regions is associated with disease. Disease associated polymorphisms may include deletion or truncation of the gene, mutations that alter expression level, that affect the activity of the protein, *etc.*

25 Changes in the promoter or enhancer sequence that may affect expression levels of aggrecanase can be compared to expression levels of the normal allele by various methods known in the art. Methods for determining promoter or enhancer strength include quantitation of the expressed natural protein; insertion of the variant control element into a vector with a reporter gene such as β -galactosidase, luciferase, chloramphenicol acetyltransferase, *etc.* that provides for
30 convenient quantitation; and the like.

A number of methods are available for analyzing nucleic acids for the presence of a specific sequence, *e.g.* a disease associated polymorphism. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express aggrecanase may be used as a source of mRNA, which may be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid may be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis. The use of the polymerase chain reaction is described in Saiki, *et al.* (1985), *Science* **239**:487, and a review of techniques may be found in Sambrook, *et al.* Molecular Cloning: A Laboratory Manual, CSH Press 1989, pp.14.2–14.33. Alternatively, various methods are known in the art that utilize oligonucleotide ligation as a means of detecting polymorphisms, for examples see Riley *et al.* (1990), *Nucl. Acids Res.* **18**:2887-2890; and Delahunty *et al.* (1996), *Am. J. Hum. Genet.* **58**:1239-1246.

A detectable label may be included in an amplification reaction. Suitable labels include fluorochromes, *e.g.* fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, *e.g.* ³²P, ³⁵S, ³H; *etc.* The label may be a two stage system, where the amplified DNA is conjugated to biotin, haptens, *etc.* having a high affinity binding partner, *e.g.* avidin, specific antibodies, *etc.*, where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

The sample nucleic acid, *e.g.* amplified or cloned fragment, is analyzed by one of a number of methods known in the art. The nucleic acid may be sequenced by dideoxy or other methods, and the sequence of bases compared to a wild-type aggrecanase gene sequence. Hybridization with the variant sequence may also be used to determine its presence, by Southern blots, dot blots, *etc.* The hybridization pattern of a control and variant sequence to an array of oligonucleotide probes immobilized on a solid support, as described in US 5,445,934, or in WO 95/35505, may also be used as a means of detecting the presence of variant sequences. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex

analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease, the sample is digested with that endonuclease, and the products size fractionated to determine whether the fragment was digested.

5 Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels.

Screening for mutations in aggrecanase may be based on the functional or antigenic characteristics of the protein. Protein truncation assays are useful in detecting deletions that may affect the biological activity of the protein. Various immunoassays designed to detect
10 polymorphisms in aggrecanase proteins may be used in screening. Where many diverse genetic mutations lead to a particular disease phenotype, functional protein assays have proven to be effective screening tools. The activity of the encoded aggrecanase protein may be determined by comparison with the wild-type protein.

Diagnostic methods of the subject invention in which the level of aggrecanase gene
15 expression is of interest will typically involve comparison of the aggrecanase nucleic acid abundance of a sample of interest with that of a control value to determine any relative differences, where the difference may be measured qualitatively and/or quantitatively, which differences are then related to the presence or absence of an abnormal aggrecanase gene expression pattern. A variety of different methods for determine the nucleic acid abundance in a sample are known to
20 those of skill in the art, where particular methods of interest include those described in: Pietu et al., Genome Res. (June 1996) 6: 492-503; Zhao et al., Gene (April 24, 1995) 156: 207-213; Soares, Curr. Opin. Biotechnol. (October 1997) 8: 542-546; Raval, J. Pharmacol Toxicol Methods (November 1994) 32: 125-127; Chalifour et al., Anal. Biochem (February 1, 1994) 216: 299-304; Stolz & Tuan, Mol. Biotechnol. (December 1996) 6: 225-230; Hong et al., Bioscience Reports
25 (1982) 2: 907; and McGraw, Anal. Biochem. (1984) 143: 298. Also of interest are the methods disclosed in WO 97/27317, the disclosure of which is herein incorporated by reference.

SCREENING ASSAYS

30 The subject aggrecanase polypeptides find use in various screening assays designed to identify therapeutic agents. In vitro screening assays can be employed in which the activity of

aggrecanase is assessed in the presence of a candidate therapeutic agent and compared to a control, i.e. the activity in the absence of the candidate therapeutic agent. Activity can be determined in a number of different ways, where activity may generally be determined as ability to cleave aggrecan or at least a fragment thereof, as well as a recombinant polypeptide, that includes the aggrecanase cleavage site, as described above. Such assays are described in U.S. Patent No. 5,872,209 and WO 99/05921, the disclosures of which are herein incorporated by reference, as well as Arner et al., J. Biol. Chem. (March 1999) 274: 6594-6601.

Also of interest in screening assays are non-human transgenic animals which express functional aggrecanase, where such animals are described above. In many embodiments, the animals lack endogenous aggrecanase. In using such animals for screening applications, a test compound(s) is administered to the animal, and the resultant changes in phenotype, e.g. presence of aggrecan produced by cleavage of the Glu³⁷³-Ala³⁷⁴ bond, are compared with a control.

Alternatively, in vitro models of aggrecanase binding activity may be measured in which binding events between aggrecanase and candidate aggrecanase modulatory agents are monitored.

A variety of other reagents may be included in the screening assays, depending on the particular screening protocols employed. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc that are used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, *etc.* may be used.

A variety of different candidate therapeutic agents that serve as either aggrecanase agonists or antagonists may be screened by the above methods. Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs.

Of particular interest in many embodiments are screening methods that identify agents that selectively modulate, e.g. inhibit, the subject aggrecanase enzyme and not other proteases.

AGGREGANASE NUCLEIC ACID AND POLYPEPTIDE THERAPEUTIC COMPOSITIONS

15

The nucleic acid compositions of the subject invention also find use as therapeutic agents in situations where one wishes to enhance aggrecanase activity in a host. The aggrecanase genes, gene fragments, or the encoded aggrecanase protein or protein fragments are useful in gene therapy to treat disorders associated with aggrecanase defects. Expression vectors may be used to introduce the aggrecanase gene into a cell. Such vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences. Transcription cassettes may be prepared comprising a transcription initiation region, the target gene or fragment thereof, and a transcriptional termination region. The transcription cassettes may be introduced into a variety of vectors, *e.g.* plasmid; retrovirus, *e.g.* lentivirus; adenovirus; and the like, where the vectors are able to transiently or stably be maintained in the cells, usually for a period of at least about one day, more usually for a period of at least about several days to several weeks.

The gene or aggrecanase protein may be introduced into tissues or host cells by any number of routes, including viral infection, microinjection, or fusion of vesicles. Jet injection may also be used for intramuscular administration, as described by Furth *et al.* (1992), *Anal Biochem* **205**:365-368. The DNA may be coated onto gold microparticles, and delivered intradermally by a particle bombardment device, or "gene gun" as described in the literature (see, for example, Tang *et al.*

(1992), *Nature* 356:152-154), where gold microprojectiles are coated with the DNA, then bombarded into skin cells.

METHODS OF MODULATING AGGRECANASE ACTIVITY

5 The subject invention provides methods of modulating aggrecanase activity in a cell, including methods of increasing aggrecanase activity (e.g. methods of enhancing), as well as methods of reducing or inhibiting aggrecanase activity, e.g. methods of stopping or limiting aggrecan cleavage. In such methods, an effective amount of an aggrecanase modulatory agent is contacted with the cell.

10 Also provided are methods of modulating, including enhancing and inhibiting, aggrecanase activity in a host. In such methods, an effective amount of active agent that modulates the activity of aggrecanase *in vivo*, e.g. usually enhances or inhibits aggrecanase activity, is administered to the host. The active agent may be a variety of different compounds, including a naturally occurring or synthetic small molecule compound, an antibody, fragment or derivative thereof, an antisense
15 composition, and the like.

Of particular interest in certain embodiments are agents that reduce aggrecanase activity, e.g. aggrecan cleavage, by at least about 10 fold, usually at least about 20 fold and more usually at least about 25 fold, as measure by the Assay described in Arner et al. (1999), *supra*. In many
20 embodiments, of particular interest is the use of compounds that reduce aggrecanase activity by at least 100 fold, as compared to a control.

Also of interest is the use of agents that, while providing for reduced aggrecanase activity, do not substantially reduce the activity of other proteinases, if at all. Thus, the agents in this embodiment are selective inhibitors of aggrecanase. An agent is considered to be selective if it provides for the above reduced aggrecanase activity, but substantially no reduced activity of at least
25 one other proteinase, where substantially no means less than 10 fold reduction, usually less than 5 fold reduction and in many instances less than 1 fold reduction, where activity is measured as described in Arner et al., (1999), *supra*.

Naturally occurring or synthetic small molecule compounds of interest include numerous chemical classes, though typically they are organic molecules, preferably small organic compounds
30 having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen

bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including
5 peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Also of interest as active agents are antibodies that at least reduce, if not inhibit, the target aggrecanase activity in the host. Suitable antibodies are obtained by immunizing a host animal with peptides comprising all or a portion of the target protein, e.g. aggrecanase. Suitable host animals
10 include mouse, rat sheep, goat, hamster, rabbit, *etc.* The origin of the protein immunogen may be mouse, human, rat, monkey *etc.* The host animal will generally be a different species than the immunogen, e.g. human aggrecanase used to immunize mice, *etc.*

The immunogen may comprise the complete protein, or fragments and derivatives thereof. Preferred immunogens comprise all or a part of aggrecanase, where these residues contain the
15 post-translation modifications, such as glycosylation, found on the native target protein. Immunogens comprising the extracellular domain are produced in a variety of ways known in the art, e.g. expression of cloned genes using conventional recombinant methods, isolation from HEC, *etc.*

For preparation of polyclonal antibodies, the first step is immunization of the host animal
20 with the target protein, where the target protein will preferably be in substantially pure form, comprising less than about 1% contaminant. The immunogen may comprise the complete target protein, fragments or derivatives thereof. To increase the immune response of the host animal, the target protein may be combined with an adjuvant, where suitable adjuvants include alum, dextran, sulfate, large polymeric anions, oil & water emulsions, e.g. Freund's adjuvant, Freund's complete
25 adjuvant, and the like. The target protein may also be conjugated to synthetic carrier proteins or synthetic antigens. A variety of hosts may be immunized to produce the polyclonal antibodies. Such hosts include rabbits, guinea pigs, rodents, e.g. mice, rats, sheep, goats, and the like. The target protein is administered to the host, usually intradermally, with an initial dosage followed by one or more, usually at least two, additional booster dosages. Following immunization, the blood from the
30 host will be collected, followed by separation of the serum from the blood cells. The Ig present in

the resultant antiserum may be further fractionated using known methods, such as ammonium salt fractionation, DEAE chromatography, and the like.

Monoclonal antibodies are produced by conventional techniques. Generally, the spleen and/or lymph nodes of an immunized host animal provide a source of plasma cells. The plasma
5 cells are immortalized by fusion with myeloma cells to produce hybridoma cells. Culture supernatant from individual hybridomas is screened using standard techniques to identify those producing antibodies with the desired specificity. Suitable animals for production of monoclonal antibodies to the human protein include mouse, rat, hamster, *etc.* To raise antibodies against the mouse protein, the animal will generally be a hamster, guinea pig, rabbit, *etc.* The antibody may be
10 purified from the hybridoma cell supernatants or ascites fluid by conventional techniques, *e.g.* affinity chromatography using aggrecanase bound to an insoluble support, protein A sepharose, *etc.*

The antibody may be produced as a single chain, instead of the normal multimeric structure. Single chain antibodies are described in Jost *et al.* (1994) J.B.C. **269**:26267-73, and others. DNA
15 sequences encoding the variable region of the heavy chain and the variable region of the light chain are ligated to a spacer encoding at least about 4 amino acids of small neutral amino acids, including glycine and/or serine. The protein encoded by this fusion allows assembly of a functional variable region that retains the specificity and affinity of the original antibody.

For *in vivo* use, particularly for injection into humans, it is desirable to decrease the antigenicity of the antibody. An immune response of a recipient against the blocking agent will
20 potentially decrease the period of time that the therapy is effective. Methods of humanizing antibodies are known in the art. The humanized antibody may be the product of an animal having transgenic human immunoglobulin constant region genes (see for example International Patent Applications WO 90/10077 and WO 90/04036). Alternatively, the antibody of interest may be engineered by recombinant DNA techniques to substitute the CH1, CH2, CH3, hinge domains,
25 and/or the framework domain with the corresponding human sequence (see WO 92/02190).

The use of Ig cDNA for construction of chimeric immunoglobulin genes is known in the art (Liu *et al.* (1987) P.N.A.S. **84**:3439 and (1987) J. Immunol. **139**:3521). mRNA is isolated from a hybridoma or other cell producing the antibody and used to produce cDNA. The cDNA of interest may be amplified by the polymerase chain reaction using specific primers (U.S. Patent nos.
30 4,683,195 and 4,683,202). Alternatively, a library is made and screened to isolate the sequence of interest. The DNA sequence encoding the variable region of the antibody is then fused to human

constant region sequences. The sequences of human constant regions genes may be found in Kabat *et al.* (1991) Sequences of Proteins of Immunological Interest, N.I.H. publication no. 91-3242.

Human C region genes are readily available from known clones. The choice of isotype will be guided by the desired effector functions, such as complement fixation, or activity in antibody-dependent cellular cytotoxicity. Preferred isotypes are IgG1, IgG3 and IgG4. Either of the human light chain constant regions, kappa or lambda, may be used. The chimeric, humanized antibody is then expressed by conventional methods.

Antibody fragments, such as Fv, F(ab')₂ and Fab may be prepared by cleavage of the intact protein, *e.g.* by protease or chemical cleavage. Alternatively, a truncated gene is designed. For example, a chimeric gene encoding a portion of the F(ab')₂ fragment would include DNA sequences encoding the CH1 domain and hinge region of the H chain, followed by a translational stop codon to yield the truncated molecule.

Consensus sequences of H and L J regions may be used to design oligonucleotides for use as primers to introduce useful restriction sites into the J region for subsequent linkage of V region segments to human C region segments. C region cDNA can be modified by site directed mutagenesis to place a restriction site at the analogous position in the human sequence.

Expression vectors include plasmids, retroviruses, YACs, EBV derived episomes, and the like. A convenient vector is one that encodes a functionally complete human CH or CL immunoglobulin sequence, with appropriate restriction sites engineered so that any VH or VL sequence can be easily inserted and expressed. In such vectors, splicing usually occurs between the splice donor site in the inserted J region and the splice acceptor site preceding the human C region, and also at the splice regions that occur within the human CH exons. Polyadenylation and transcription termination occur at native chromosomal sites downstream of the coding regions. The resulting chimeric antibody may be joined to any strong promoter, including retroviral LTRs, *e.g.* SV-40 early promoter, (Okayama *et al.* (1983) Mol. Cell. Bio. 3:280), Rous sarcoma virus LTR (Gorman *et al.* (1982) P.N.A.S. 79:6777), and moloney murine leukemia virus LTR (Grosschedl *et al.* (1985) Cell 41:885); native Ig promoters, *etc.*

In yet other embodiments of the invention, the active agent is an agent that modulates, and generally decreases or down regulates, the expression of the gene encoding the target protein in the host. For example, antisense molecules can be used to down-regulate expression of aggrecanase in

cells. The anti-sense reagent may be antisense oligonucleotides (ODN), particularly synthetic ODN having chemical modifications from native nucleic acids, or nucleic acid constructs that express such anti-sense molecules as RNA. The antisense sequence is complementary to the mRNA of the targeted gene, and inhibits expression of the targeted gene products. Antisense molecules inhibit gene expression through various mechanisms, *e.g.* by reducing the amount of mRNA available for translation, through activation of RNase H, or steric hindrance. One or a combination of antisense molecules may be administered, where a combination may comprise multiple different sequences.

Antisense molecules may be produced by expression of all or a part of the target gene sequence in an appropriate vector, where the transcriptional initiation is oriented such that an antisense strand is produced as an RNA molecule. Alternatively, the antisense molecule is a synthetic oligonucleotide. Antisense oligonucleotides will generally be at least about 7, usually at least about 12, more usually at least about 20 nucleotides in length, and not more than about 500, usually not more than about 50, more usually not more than about 35 nucleotides in length, where the length is governed by efficiency of inhibition, specificity, including absence of cross-reactivity, and the like. It has been found that short oligonucleotides, of from 7 to 8 bases in length, can be strong and selective inhibitors of gene expression (see Wagner *et al.* (1996), *Nature Biotechnol.* 14:840-844).

A specific region or regions of the endogenous sense strand mRNA sequence is chosen to be complemented by the antisense sequence. Selection of a specific sequence for the oligonucleotide may use an empirical method, where several candidate sequences are assayed for inhibition of expression of the target gene in an *in vitro* or animal model. A combination of sequences may also be used, where several regions of the mRNA sequence are selected for antisense complementation.

Antisense oligonucleotides may be chemically synthesized by methods known in the art (see Wagner *et al.* (1993), *supra*, and Milligan *et al.*, *supra*.) Preferred oligonucleotides are chemically modified from the native phosphodiester structure, in order to increase their intracellular stability and binding affinity. A number of such modifications have been described in the literature, which alter the chemistry of the backbone, sugars or heterocyclic bases.

Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulfur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O'-5'-S-phosphorothioate, 3'-S-5'-O-phosphorothioate, 3'-CH₂-5'-O-phosphonate and

3'-NH-5'-O-phosphoroamidate. Peptide nucleic acids replace the entire ribose phosphodiester backbone with a peptide linkage. Sugar modifications are also used to enhance stability and affinity. The α -anomer of deoxyribose may be used, where the base is inverted with respect to the natural β -anomer. The 2'-OH of the ribose sugar may be altered to form 2'-O-methyl or 2'-O-allyl sugars, which provides resistance to degradation without comprising affinity. Modification of the heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2'-deoxycytidine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5-propynyl-2'-deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological activity when substituted for deoxythymidine and deoxycytidine, respectively.

As an alternative to anti-sense inhibitors, catalytic nucleic acid compounds, *e.g.* ribozymes, anti-sense conjugates, *etc.* may be used to inhibit gene expression. Ribozymes may be synthesized *in vitro* and administered to the patient, or may be encoded on an expression vector, from which the ribozyme is synthesized in the targeted cell (for example, see International patent application WO 9523225, and Beigelman *et al.* (1995), *Nucl. Acids Res.* **23**:4434-42). Examples of oligonucleotides with catalytic activity are described in WO 9506764. Conjugates of anti-sense ODN with a metal complex, *e.g.* terpyridylCu(II), capable of mediating mRNA hydrolysis are described in Bashkin *et al.* (1995), *Appl. Biochem. Biotechnol.* **54**:43-56.

As mentioned above, an effective amount of the active agent is administered to the host, where "effective amount" means a dosage sufficient to produce a desired result. Generally, the desired result is at least an enhancement or reduction in aggrecanase activity, as measured by aggrecan cleavage product production, as compared to a control.

In the subject methods, the active agent(s) may be administered to the host using any convenient means capable of resulting in the desired modulation of aggrecanase activity, *e.g.* desired reduction in aggrecan cleavage product production. Thus, the agent can be incorporated into a variety of formulations for therapeutic administration. More particularly, the agents of the present invention can be formulated into pharmaceutical compositions by combination with appropriate, pharmaceutically acceptable carriers or diluents, and may be formulated into preparations in solid, semi-solid, liquid or gaseous forms, such as tablets, capsules, powders, granules, ointments, solutions, suppositories, injections, inhalants and aerosols.

As such, administration of the agents can be achieved in various ways, including oral, buccal, rectal, parenteral, intraperitoneal, intradermal, transdermal, intracheal, etc., administration.

In pharmaceutical dosage forms, the agents may be administered in the form of their pharmaceutically acceptable salts, or they may also be used alone or in appropriate association, as well as in combination, with other pharmaceutically active compounds. The following methods and excipients are merely exemplary and are in no way limiting.

For oral preparations, the agents can be used alone or in combination with appropriate additives to make tablets, powders, granules or capsules, for example, with conventional additives, such as lactose, mannitol, corn starch or potato starch; with binders, such as crystalline cellulose, cellulose derivatives, acacia, corn starch or gelatins; with disintegrators, such as corn starch, potato starch or sodium carboxymethylcellulose; with lubricants, such as talc or magnesium stearate; and if desired, with diluents, buffering agents, moistening agents, preservatives and flavoring agents.

The agents can be formulated into preparations for injection by dissolving, suspending or emulsifying them in an aqueous or nonaqueous solvent, such as vegetable or other similar oils, synthetic aliphatic acid glycerides, esters of higher aliphatic acids or propylene glycol; and if desired, with conventional additives such as solubilizers, isotonic agents, suspending agents, emulsifying agents, stabilizers and preservatives.

The agents can be utilized in aerosol formulation to be administered via inhalation. The compounds of the present invention can be formulated into pressurized acceptable propellants such as dichlorodifluoromethane, propane, nitrogen and the like.

Furthermore, the agents can be made into suppositories by mixing with a variety of bases such as emulsifying bases or water-soluble bases. The compounds of the present invention can be administered rectally via a suppository. The suppository can include vehicles such as cocoa butter, carbowaxes and polyethylene glycols, which melt at body temperature, yet are solidified at room temperature.

Unit dosage forms for oral or rectal administration such as syrups, elixirs, and suspensions may be provided wherein each dosage unit, for example, teaspoonful, tablespoonful, tablet or suppository, contains a predetermined amount of the composition containing one or more inhibitors. Similarly, unit dosage forms for injection or intravenous administration may comprise the inhibitor(s) in a composition as a solution in sterile water, normal saline or another pharmaceutically acceptable carrier.

The term "unit dosage form," as used herein, refers to physically discrete units suitable as unitary dosages for human and animal subjects, each unit containing a predetermined quantity of compounds of the present invention calculated in an amount sufficient to produce the desired effect in association with a pharmaceutically acceptable diluent, carrier or vehicle. The specifications for the novel unit dosage forms of the present invention depend on the particular compound employed and the effect to be achieved, and the pharmacodynamics associated with each compound in the host.

The pharmaceutically acceptable excipients, such as vehicles, adjuvants, carriers or diluents, are readily available to the public. Moreover, pharmaceutically acceptable auxiliary substances, such as pH adjusting and buffering agents, tonicity adjusting agents, stabilizers, wetting agents and the like, are readily available to the public.

Where the agent is a polypeptide, polynucleotide, analog or mimetic thereof, e.g. antisense composition, it may be introduced into tissues or host cells by any number of routes, including viral infection, microinjection, or fusion of vesicles. Jet injection may also be used for intramuscular administration, as described by Furth *et al.* (1992), *Anal Biochem* **205**:365-368. The DNA may be coated onto gold microparticles, and delivered intradermally by a particle bombardment device, or "gene gun" as described in the literature (see, for example, Tang *et al.* (1992), *Nature* **356**:152-154), where gold microprojectiles are coated with the therapeutic DNA, then bombarded into skin cells.

Those of skill in the art will readily appreciate that dose levels can vary as a function of the specific compound, the severity of the symptoms and the susceptibility of the subject to side effects. Preferred dosages for a given compound are readily determinable by those of skill in the art by a variety of means.

The subject methods find use in the treatment of a variety of different disease conditions involving aggrecanase activity. Of particular interest is the use of the subject methods to treat disease conditions characterized by the presence of aggrecan cleavage products, particularly 60 kDa aggrecan cleavage products having an ARGS N-terminus. Specific diseases that are characterized by the presence of such methods include: rheumatoid arthritis, osteo-arthritis, infectious arthritis, gouty arthritis, psoriatic arthritis, spondylitis, sports injury, joint trauma, pulmonary disease, fibrosis, and the like.

By treatment is meant at least an amelioration of the symptoms associated with the pathological condition afflicting the host, where amelioration is used in a broad sense to refer to at

least a reduction in the magnitude of a parameter, *e.g.* symptom, associated with the pathological condition being treated, such as hyperphosphatemia. As such, treatment also includes situations where the pathological condition, or at least symptoms associated therewith, are completely inhibited, *e.g.* prevented from happening, or stopped, *e.g.* terminated, such that the host no longer suffers from the pathological condition, or at least the symptoms that characterize the pathological condition.

A variety of hosts are treatable according to the subject methods. Generally such hosts are "mammals" or "mammalian," where these terms are used broadly to describe organisms which are within the class mammalia, including the orders carnivore (*e.g.*, dogs and cats), rodentia (*e.g.*, mice, guinea pigs, and rats), and primates (*e.g.*, humans, chimpanzees, and monkeys). In many embodiments, the hosts will be humans.

Kits with unit doses of the active agent, usually in oral or injectable doses, are provided. In such kits, in addition to the containers containing the unit doses will be an informational package insert describing the use and attendant benefits of the drugs in treating pathological condition of interest. Preferred compounds and unit doses are those described herein above.

The following examples are offered primarily for purposes of illustration. It will be readily apparent to those skilled in the art that the formulations, dosages, methods of administration, and other parameters of this invention may be further modified or substituted in various ways without departing from the spirit and scope of the invention.

EXPERIMENTAL

A nucleic acid array carrying 699 known metalloproteinase genes and novel ESTs available in public and proprietary databases was designed. These sequences on the array were selected by a search with a seed set of known metalloprotease protein sequences from all species. These protein sequences were used to find matching sequences in human nucleotide databases (GenBank and LifeSeq) at the protein (codon) level. Redundant sequences were eliminated, remaining sequences assembled and clustered, and the unique set of 699 sequences were arrayed.

The resultant array was used to screen genes expressed in primary cultures of synoviocytes and chondrocytes. A fair number of metalloproteinases known to be expressed by these cells were identified. However, a number of ESTs for novel proteins were also identified. One of these ESTs was MPS5584x1. Upon further examination, this EST was found to be the human homolog of the murine ADAMTS-1, which has previously been cloned.

From the Incyte database, by homology search, we obtained other colinear human ESTs corresponding to the previously cloned mouse sequence. By synthesizing corresponding primers, the entire human ADAMTS cDNA was synthesized using PCR protocols. Figure 1 provides the full length sequence of the human ADAMTS-cDNA (also referred to herein as human aggrecanase).

An alignment analysis of this ADAMTS cDNA with the two ADMPs reported in WO 99/05291 (seq1 and seq2) and the novel metalloprotease reported in WO 97/31931 (w35293) was done using GAP (Needleman-Wunsch). (GAP program in GCG 10.0. A pairwise DNA alignment was made between each of the sequences; scoring matrix = PAM250. Gap and BestFit were originally written for Version 1.0 by Paul Haeberli from a careful reading of the Needleman and Wunsch (J. Mol. Biol. 48; 443-453 (1970)) and the Smith and Waterman (Adv. Appl. Math. 2; 482-489 (1981)) papers. Limited alignments were designed by Paul Haeberli and added to the Package for Versions 3.0. They were united into a single program by Philip Delaquess for Version 4.0).

The following results were obtained:

Pairwise comparisons using Gap (Needleman-Wunsch),
Matrix PAM250.

	seq1	seq2	w35293	
adamts1hum	50%	49%	33%	(similarity)
	38%	39%	15%	(identity)
seq1		43%	28%	
		32%	19%	
seq2			35%	
			18%	

The human ADAMTS is clearly different from seq1, seq2 and w35293.

It is apparent from the above results and discussion that human aggrecanase, as well as polypeptides related thereto and nucleic acid compositions encoding the same, are provided by the subject invention. These polypeptide and nucleic acid compositions find use in a variety of diverse applications, including research, diagnostic, screening and therapeutic applications. Also provided are novel methods of treating diseases associated aggrecanase, as the identification of the subject aggrecanase provides for an additional target for therapeutic agents for such diseases. Accordingly, the subject invention provides for a significant contribution to the field.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.